

An Augmented Lagrangian Coordination-Decomposition Algorithm for Solving Distributed Non-Convex Programs

Jean-Hubert Hours and Colin N. Jones

Abstract—A novel augmented Lagrangian method for solving non-convex programs with nonlinear cost and constraint couplings in a distributed framework is presented. The proposed decomposition algorithm is made of two layers: The outer level is a standard multiplier method with penalty on the nonlinear equality constraints, while the inner level consists of a block-coordinate descent (BCD) scheme. Based on standard results on multiplier methods and recent results on proximal regularised BCD techniques, it is proven that the method converges to a KKT point of the non-convex nonlinear program under a semi-algebraicity assumption. Efficacy of the algorithm is demonstrated on a numerical example.

I. INTRODUCTION

When dealing with a large-scale system, such as a power grid for instance, sub-systems are coupled in a complex manner through their dynamics, implying that local control actions may have a major impact throughout the whole network. Implementing NMPC controllers for such systems may result in large-scale non-separable non-convex nonlinear programs (NLP). Solving such programs online in a centralised fashion is a challenging task from a computational point of view. Moreover, autonomy of the agents is likely to be hampered by such a centralised procedure. Therefore, distributed MPC is currently raising much interest, as an advanced control strategy for decentralising computations, thus reducing the computational burden and enabling autonomy of agents, which then only need to solve a local NLP [8]. Potential applications of distributed NMPC abound, from networked systems' control (interconnected power plants) to collaborative control (formation flying). Decomposing a convex NLP is generally performed by applying Lagrangian decomposition methods [4]. A critical requirement for this strategy is strong duality, which is guaranteed by Slater's condition in the convex case. However, such an assumption rarely holds in a non-convex setting. Nevertheless, in an augmented Lagrangian setting, it has been shown that the duality gap can be locally removed by using appropriate penalty functions [10]. Moreover, augmented Lagrangian methods turn out to be much more efficient than standard penalty approaches from a computational point of view, as the risk of ill-conditioning associated with large penalty parameters is reduced by the fast convergence of the dual iteration. However, the resulting quadratic penalty term is not separable even if the coupling constraints are. Several approaches have been explored to remedy this issue [7]. For instance, in [2,12],

a local convexification procedure by means of proximal regularisation is analysed. A more recent approach to the decomposition of non-convex programs is the sequential programming scheme presented in [9], which consists in iterative convex approximations and decompositions. Our decomposition strategy is not specifically targeted at separable NLPs, as objective and constraints couplings can be addressed by the proposed BCD scheme. It also differs from [9] in that only local convexification is required to guarantee convergence.

Our approach is an augmented Lagrangian method with partial constraint penalisation on the nonlinear equality constraints. It is a two-level optimisation scheme, whose outer layer is a loop on the Lagrange multipliers estimates associated with the nonlinear equality constraints. The inner level consists in approximately solving the primal problem to a critical point. This is performed via an inexact proximal regularised BCD scheme. By means of recent results on non-convex proximal regularised BCD methods [1], we prove convergence of the inner loop to a critical point of the non-convex inner problem under the assumption that the NLP is semi-algebraic. This is the main novelty of our approach, since, until recently, BCD type methods were thought to only have convergence guarantees under the very restrictive assumption that the NLP is convex. Some studies have been conducted in the non-convex case, yet the convergence results are quite weak compared to [1]. The inner loop is the key step of the algorithm, which enables distributed computations. In general, as the augmented Lagrangian is not separable, inner iterations cannot be fully parallelised, but some degree of parallelism can be achieved if the interconnection graph is sparse [4]. Finally, convergence of our two-level optimisation technique to a KKT point of the original NLP is proven under standard assumptions. Eventually, the whole algorithm consists in iteratively solving decoupled Quadratic Programs (QP) and updating multipliers estimates in a coordination step. In Section III, the general framework of our algorithm is presented. In Sections IV and V, the algorithm is described and its convergence analysed. Finally, a numerical example is presented in Section VI.

II. BACKGROUND

Definition 1 (Normal cone to a convex set): Let Ω be a convex set in \mathbb{R}^n and $\bar{x} \in \Omega$. The normal cone to Ω at \bar{x} is the set

$$\mathcal{N}_{\Omega}(\bar{x}) := \left\{ v \in \mathbb{R}^n \mid \forall x \in \Omega, v^{\top}(x - \bar{x}) \leq 0 \right\}. \quad (1)$$

Jean-Hubert Hours and Colin N. Jones are with the Laboratoire d'Automatique, École Polytechnique Fédérale de Lausanne, Switzerland. {jean-hubert.hours, colin.jones}@epfl.ch

Lemma 1 (Descent Lemma): Let $f \in C^2(\Omega, \mathbb{R})$ (twice continuously differentiable in Ω), where Ω is a convex compact set in \mathbb{R}^n . Let $x, y \in \Omega$.

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{\|\nabla^2 f\|_\infty^\Omega}{2} \|y - x\|_2^2 \quad (2)$$

where

$$\|\nabla^2 f\|_\infty^\Omega = \max \left\{ \|\nabla^2 f(x)\|_2 \mid x \in \Omega \right\}. \quad (3)$$

Given $M \in \mathbb{R}^{n \times n}$, $\|M\|_2$ denotes the induced matrix 2-norm.

Definition 2 (Regular point): Let $h \in C^1(\mathbb{R}^n, \mathbb{R}^m)$. A point $x \in \mathbb{R}^n$ such that $h(x) = 0$ is called *regular* if the gradients $\nabla h_1(x), \dots, \nabla h_m(x)$ are linearly independent.

Definition 3 (Critical point): Let f be a proper lower semicontinuous function. A necessary condition for x^* to be a minimiser of f is that

$$0 \in \partial f(x^*), \quad (4)$$

where $\partial f(x^*)$ is the sub-differential of f at x^* [11]. Points satisfying (4) are called *critical points*.

The indicator function of a closed subset S of \mathbb{R}^n is denoted δ_S and is defined as

$$\delta_S(x) = \begin{cases} 0 & \text{if } x \in S \\ +\infty & \text{if } x \notin S. \end{cases} \quad (5)$$

Lemma 2 (Sub-differential of indicator function [11]):

Given a convex set C , for all $x \in C$,

$$\partial \delta_C(x) = \mathcal{N}_C(x). \quad (6)$$

III. PROBLEM FORMULATION

We consider NLPs of the following form:

$$\begin{aligned} & \underset{z_1, \dots, z_N}{\text{minimise}} \quad \sum_{i=1}^N J_i(z_i) + Q(z_1, \dots, z_N) \\ & \text{s.t.} \quad F_i(z_i) = 0, \quad i \in \{1, \dots, N\}, \\ & \quad G(z_1, \dots, z_N) = 0, \\ & \quad z_i \in \mathcal{Z}_i, \quad i \in \{1, \dots, N\}, \end{aligned} \quad (7)$$

where $z_i \in \mathbb{R}^{n_i}$ for $i \in \{1, \dots, N\}$ and \mathcal{Z}_i are polytopes in \mathbb{R}^{n_i} . The term $Q(z_1, \dots, z_N)$ is a *cost coupling* term and $G(z_1, \dots, z_N) \in \mathbb{R}^p$ a *constraint coupling* term. They model the different kinds of systems' interactions, which may appear in a distributed NMPC context.

For the remainder of the paper, we also define the vector $z := (z_1^\top, \dots, z_N^\top)^\top \in \mathbb{R}^n$, with dimension $n := \sum_{i=1}^N n_i$. For $i \in \{1, \dots, N\}$, as \mathcal{Z}_i is polytopic, there exists $A_i \in \mathbb{R}^{q_i \times n_i}$ and $b_i \in \mathbb{R}^{q_i}$ such that $\mathcal{Z}_i := \{x \in \mathbb{R}^{n_i} \mid A_i x \leq b_i\}$. One can thus introduce the polytope $\mathcal{Z} \subset \mathbb{R}^n$ by $\mathcal{Z} := \{x \in \mathbb{R}^n \mid Ax \leq b\}$, where $A \in \mathbb{R}^{q \times n}$, $b \in \mathbb{R}^q$ and $q := \sum_{i=1}^N q_i$. By posing $m := \sum_{i=1}^N m_i$, we also define

$$H(z) := (F_1(z_1)^\top, \dots, F_N(z_N)^\top, G(z_1, \dots, z_N)^\top)^\top \in \mathbb{R}^r, \quad (8)$$

where $r := m + p$. The distributed objective is

$$J(z) := \sum_{i=1}^N J_i(z_i) + Q(z_1, \dots, z_N), \quad (9)$$

so that NLP (7) can be rewritten

$$\begin{aligned} & \underset{z}{\text{minimise}} \quad J(z) \\ & \text{s.t.} \quad H(z) = 0, \quad z \in \mathcal{Z}. \end{aligned} \quad (10)$$

Assumption 1 (Smoothness and semi-algebraicity):

The functions $\{J_i\}_{i=1}^N$, Q , $\{F_i\}_{i=1}^N$ and G are twice continuously differentiable and semi-algebraic.

Remark 1: The set of semi-algebraic functions is closed with respect to sum, product and composition. The indicator function of a semi-algebraic set is a semi-algebraic function. Semi-algebraicity is needed in order to apply the results of [1], which are valid for functions satisfying the Kurdyka-Lojasiewicz (KL) property, which is the case for all real semi-algebraic functions [5]. From a control perspective, this means that our results are valid for polynomial systems subject to polynomial constraints and objectives.

Assumption 2: The NLP (10) admits an isolated KKT point $((z^*)^\top, (\mu^*)^\top, (\lambda^*)^\top)^\top \in \mathbb{R}^{n+r+q}$, which satisfies

- z^* is regular,
- $((z^*)^\top, (\mu^*)^\top, (\lambda^*)^\top)^\top$ satisfies the *second order optimality condition*, that is

$$\begin{aligned} & p^\top \nabla^2 L(z^*, \mu^*, \lambda^*) p > 0 \text{ for all } p \in \mathbb{R}^n \text{ such that} \\ & \nabla H(z^*)^\top p = 0, \quad A_{I^*} p = 0, \end{aligned} \quad (11)$$

where $I^* \subseteq \{1, \dots, q\}$ is the set of indices of active inequality constraints at z^* .

- For all $i \in I^*$, $\lambda_i^* > 0$.

IV. OUTER LOOP: PARTIALLY AUGMENTED LAGRANGIAN

Given a penalty parameter $\rho > 0$, the augmented Lagrangian associated with problem (10) is defined as

$$L_\rho(z, \mu) := J(z) + \mu^\top H(z) + \frac{\rho}{2} \|H(z)\|_2^2, \quad (12)$$

where $\mu \in \mathbb{R}^r$. Only equality constraints $H(z) = 0$ are penalised. The polytopic inequalities are kept as constraints on the augmented Lagrangian.

A. Algorithm description

The outer loop is similar to a dual ascent on the multiplier estimates μ_k , along with iterative updates of the penalty parameter ρ_k . At each outer iteration k , the primal problem

$$\underset{z \in \mathcal{Z}}{\text{minimise}} \quad L_{\rho_k}(z, \mu_k) \quad (13)$$

is solved to a given level of accuracy $\epsilon_k > 0$. The outer loop is presented in Algorithm 1 below. The notation $d(x, S)$ stands for the distance function between a point x and a set S in \mathbb{R}^n and is defined by

$$d(x, S) := \inf_{z \in S} \|x - z\|_2. \quad (14)$$

At every outer iteration, a point \bar{z} is found, which satisfies inexact optimality conditions. Then, the dual estimate μ is

Algorithm 1 Method of multipliers with partial constraint penalisation

Input: Objective J , equality constraint H , polytope \mathcal{Z} , $\beta > 1$ and final tolerance on nonlinear equality constraints $\eta > 0$.

- Initial guess for optimiser $z_0 \in \mathcal{Z}$, initial guess for multiplier estimates $\mu_0 \in \mathbb{R}^r$, initial penalty parameter $\rho_0 > 1$, initial tolerance on optimality conditions $\epsilon_0 > 0$.

Initialization: $z \leftarrow z_0$, $\mu \leftarrow \mu_0$, $\rho \leftarrow \rho_0$, $\epsilon \leftarrow \epsilon_0$.

repeat

Find $\bar{z} \in \mathcal{Z}$ such that $d(0, \nabla L_\rho(\bar{z}, \mu) + \mathcal{N}_{\mathcal{Z}}(\bar{z})) \leq \epsilon$
using z as a warm-start

$z \leftarrow \bar{z}$

$\mu \leftarrow \mu + \rho H(z)$, $\epsilon \leftarrow \epsilon/\rho$, $\rho \leftarrow \beta\rho$

until $\|H(z)\| \leq \eta$

updated, the penalty parameter increased and the tolerance on the first order optimality conditions is reduced. The main tuning variables are the initial value of the penalty parameter ρ_0 and the growth coefficient β .

B. Convergence analysis

We prove that the augmented Lagrangian iterations are locally convergent to the isolated KKT point $((z^*)^\top, (\mu^*)^\top, (\lambda^*)^\top)^\top$ satisfying Assumption 2. Our analysis is based on the results of [3]. We start by showing that the inner problem is well-defined, which means that for any optimality tolerance $\epsilon > 0$, there exists a point \bar{z} , as in Algorithm 1, satisfying $d(0, \nabla L_\rho(\bar{z}, \mu) + \mathcal{N}_{\mathcal{Z}}(\bar{z})) \leq \epsilon$, under appropriate conditions on ρ and μ .

Lemma 3 (Existence of an inner critical point): There exists $\delta > 0$, $\bar{\rho} > 0$ and $\kappa > 0$ such that

$$\forall (\mu, \rho) \in \mathcal{S} := \{(\mu, \rho) \in \mathbb{R}^{m+1} \mid \|\mu - \mu^*\|_2 \leq \delta\rho, \rho \geq \bar{\rho}\}, \quad (15)$$

the problem

$$\begin{aligned} &\text{minimise } L_\rho(z, \mu) \\ &\text{s.t. } z \in \mathcal{Z} \end{aligned} \quad (16)$$

$$\|z - z^*\|_2 < \kappa.$$

has a unique minimiser. Moreover, for all $\epsilon > 0$ and for all $(\mu, \rho) \in \mathcal{S}$, there exists $z_\epsilon \in \mathbb{R}^n$ such that

$$d(0, \nabla L_\rho(z_\epsilon, \mu) + \mathcal{N}_{\mathcal{Z}}(z_\epsilon)) \leq \epsilon. \quad (17)$$

Proof: The first part of the statement is a direct consequence of Proposition 2.11 in [3], as the second order optimality condition is satisfied (Assumption 2). From Definition 3, taking $(\mu, \rho) \in \mathcal{S}$, this implies that we can find $\tilde{z} \in \mathcal{Z} \cap \mathcal{B}(z^*, \kappa)$ such that $d(0, \nabla L_\rho(\tilde{z}, \mu) + \mathcal{N}_{\mathcal{Z} \cap \mathcal{B}(z^*, \kappa)}(\tilde{z})) \leq \epsilon$ for all $\epsilon > 0$, where $\mathcal{B}(z^*, \kappa)$ is the open ball of radius κ centered at z^* . As $\text{ri } \mathcal{B}(z^*, \kappa) \cap \text{ri } \mathcal{Z} \neq \emptyset$, where ri stands for the relative interior, $\mathcal{N}_{\mathcal{Z} \cap \mathcal{B}(z^*, \kappa)}(\tilde{z}) = \mathcal{N}_{\mathcal{Z}}(\tilde{z}) \cap \mathcal{N}_{\mathcal{B}(z^*, \kappa)}(\tilde{z})$, so that

$$\begin{aligned} d(0, \nabla L_\rho(\tilde{z}, \mu) + \mathcal{N}_{\mathcal{Z}}(\tilde{z})) &\leq d(0, \nabla L_\rho(\tilde{z}, \mu) \\ &\quad + \mathcal{N}_{\mathcal{Z} \cap \mathcal{B}(z^*, \kappa)}(\tilde{z})) \leq \epsilon. \end{aligned} \quad (18)$$

The end of the proof follows by taking $z_\epsilon := \tilde{z}$. ■

The next Lemma is a reformulation of the inexact optimality conditions $d(0, \nabla L_\rho(\bar{z}, \mu) + \mathcal{N}_{\mathcal{Z}}(\bar{z})) \leq \epsilon$.

Lemma 4 (Inexact optimality conditions): Let $z \in \mathbb{R}^n$, $f \in C^1(\mathbb{R}^n, \mathbb{R})$ and \mathcal{Z} be a convex set in \mathbb{R}^n . The following equivalence holds:

$$d(0, \nabla f(z) + \mathcal{N}_{\mathcal{Z}}(z)) \leq \epsilon \Leftrightarrow \exists v \in \mathbb{R}^n \text{ such that}$$

$$\begin{cases} \|v\|_2 \leq \epsilon \\ -\nabla f(z) \in \mathcal{N}_{\mathcal{Z}}(z) + v. \end{cases} \quad (19)$$

Proof: This directly follows from the definition of the distance as an infimum. ■

As $((z^*)^\top, (\mu^*)^\top, (\lambda^*)^\top)^\top$ is an isolated KKT point, there exists $\nu > 0$ such that z^* is the only critical point of (10) in $\mathcal{B}(z^*, \nu)$. From the statement of Algorithm 1, the penalty parameter ρ is guaranteed to increase to infinity and the optimality tolerance ϵ to converge to zero. The following convergence theorem is valid if the iterates z_k stay in the ball $\mathcal{B}(z^*, \min\{\nu, \kappa\})$ for a large enough k . As noticed in [3], it is generally the case in practice, as warm-starting the inner problem (16) on the previous solution leads the iterates to stay around the same local minimum z^* . Thus one can reasonably assume that $z_k \in \mathcal{B}(z^*, \min\{\nu, \kappa\})$.

Theorem 1 (Local convergence to a KKT point): Let $\{z_k\}$ and $\{\mu_k\}$ be sequences in \mathbb{R}^n such that

$$\begin{cases} z_k \in \mathcal{Z} \cap \text{cl}(\mathcal{B}(z^*, \min\{\nu, \kappa\})) \\ -\nabla L_{\rho_k}(z_k, \mu_k) \in \mathcal{N}_{\mathcal{Z}}(z_k) + d_k \end{cases} \quad (20)$$

where

- $\{\rho_k\}$ is increasing and $\rho_k \rightarrow +\infty$,
- $\{\mu_k\}$ and $\{\lambda_k\}$, associated with $\mathcal{N}_{\mathcal{Z}}(z_k)$, are bounded,
- $\|d_k\|_2 \rightarrow 0$.

Assume that all limit points of $\{z_k\}$ are regular. We then have that $z_k \rightarrow z^*$ and $\tilde{\mu}_k \rightarrow \mu^*$, where $\tilde{\mu}_k := \mu_k + \rho_k H(z_k)$, z^* and μ^* are defined in Assumption 2.

Proof: As $\mathcal{Z} \cap \text{cl}(\mathcal{B}(z^*, \min\{\nu, \kappa\}))$ is compact, by Weierstrass theorem, there exists an increasing mapping $\phi : \mathbb{N} \rightarrow \mathbb{N}$ and a point $z' \in \mathcal{Z} \cap \text{cl}(\mathcal{B}(z^*, \min\{\nu, \kappa\}))$ such that $\{z_{\phi(k)}\}$ converges to z' , which is regular by assumption. We also know by assumption that

$$\begin{aligned} \exists v_{\phi(k)} \in \mathcal{N}_{\mathcal{Z}}(z_{\phi(k)}), \\ 0 = \nabla L_{\rho_{\phi(k)}}(z_{\phi(k)}, \mu_{\phi(k)}) + v_{\phi(k)} + d_{\phi(k)}. \end{aligned} \quad (21)$$

By definition of the normal cone to \mathcal{Z} at $z_{\phi(k)}$, this means that

$$\exists \lambda_{\phi(k)} \geq 0, \begin{cases} \nabla L_{\rho_{\phi(k)}}(z_{\phi(k)}, \mu_{\phi(k)}) + A^\top \lambda_{\phi(k)} + d_{\phi(k)} = 0 \\ \lambda_{\phi(k)} \geq 0, \lambda_{\phi(k)}^\top (b - Az_{\phi(k)}) = 0. \end{cases} \quad (22)$$

We thus have

$$\nabla J(z_{\phi(k)}) + \nabla H(z_{\phi(k)})^\top \tilde{\mu}_{\phi(k)} + A^\top \lambda_{\phi(k)} = -d_{\phi(k)}. \quad (23)$$

As z' is regular, there exists $K \in \mathbb{N}_+$ such that $\nabla H(z_{\phi(k)})$ is full-rank for all $k \geq K$. Subsequently, for all $k \geq K$,

$$\begin{aligned} \tilde{\mu}_{\phi(k)} &= (\nabla H(z_{\phi(k)}) \nabla H(z_{\phi(k)})^\top)^{-1} \nabla H(z_{\phi(k)}) (-d_{\phi(k)} \\ &\quad - \nabla J(z_{\phi(k)}) - A^\top \lambda_{\phi(k)}). \end{aligned} \quad (24)$$

As $z_{\phi(k)} \rightarrow z'$, by continuity of $\lambda_{\phi(k)}^\top (b - Az_{\phi(k)}) = 0$, there exists $\lambda' \geq 0$ such that

$$\lambda_{\phi(k)} \rightarrow \lambda', (\lambda')^\top (b - Az') = 0. \quad (25)$$

As $d_{\phi(k)} \rightarrow 0$, by continuity of ∇H , this implies that $\tilde{\mu}_{\phi(k)} \rightarrow \tilde{\mu}'$, where

$$\tilde{\mu}' := (\nabla H(z') \nabla H(z')^\top)^{-1} \nabla H(z') (-\nabla J(z') - A^\top \lambda'). \quad (26)$$

By taking limit in (23), we obtain

$$\begin{cases} \nabla J(z') + \nabla H(z')^\top \tilde{\mu}' + A^\top \lambda' = 0 \\ \lambda' \geq 0, (\lambda')^\top (b - Az') = 0. \end{cases} \quad (27)$$

Feasibility of z' immediately follows from the convergence of $\tilde{\mu}_{\phi(k)}$ and the fact that $\rho_{\phi(k)} \rightarrow +\infty$. Thus we have $H(z') = 0$. This implies that $((z')^\top, ((\tilde{\mu}')^\top, ((\lambda')^\top)^\top)$ satisfies the KKT conditions.

As $z' \in \mathcal{Z} \cap \text{cl } \mathcal{B}(z^*, \min\{\kappa, \nu\})$, in which z^* is the unique KKT point by assumption, one can claim that $z' = z^*$. Taking the KKT conditions on z^* and z' , it follows that

$$\nabla H(z^*)^\top \tilde{\mu}^* + A_{I^*}^\top \lambda_{I^*}^* = \nabla H(z')^\top \tilde{\mu}' + A_{I'}^\top \lambda_{I'}', \quad (28)$$

where I^* and I' are the sets of indices of active constraints at z^* and z' respectively. Obviously, as $z^* = z'$, $I^* = I'$. Thus

$$\nabla H(z^*)^\top (\tilde{\mu}^* - \tilde{\mu}') + A_{I^*}^\top (\lambda_{I^*}^* - \lambda_{I'}') = 0. \quad (29)$$

As z^* is regular, we finally obtain that $\tilde{\mu}' = \tilde{\mu}^*$. As all limit points of $\{z_k\}$ converge to z^* , one can conclude that $z_k \rightarrow z^*$ and $\tilde{\mu}_k \rightarrow \mu^*$. ■

Remark 2: The convergence result of Theorem 1 is local. Yet convergence can be globalised (meaning that the assumption according to which the iterates lie in a ball centered at z^* can be removed) by applying the dual update scheme proposed in [6].

V. INNER LOOP: INEXACT PROXIMAL REGULARISED BCD

The non-convex inner problem (13) is solved in a distributed manner. An inexact BCD scheme is proposed, which is based on the abstract result of [1]. The main idea is to perform local convex approximations of each agent's cost and compute descent updates from it. Contrary to the approach of [9], where splitting is, in some sense, applied after convexification, we show that convexification can be performed after splitting, under the assumption that the KL property is satisfied.

A. Algorithm description

The proposed algorithm is a proximal regularised inexact BCD, which is based on Theorem 6.2 in [1], providing general properties ensuring global convergence to a critical point of the nonsmooth objective, assuming that it satisfied the KL property. The inner minimisation method is presented in Algorithm 2 below. The positive real numbers $\{\alpha_i^l\}_{i=1}^N$ are

the regularisation parameters. We assume that

$$\begin{aligned} \forall i \in \{1, \dots, N\}, \exists \alpha_i^-, \alpha_i^+ > 0 \text{ such that} \\ \begin{cases} \alpha_i^- < \alpha_i^+ \\ \forall l \in \mathbb{N}, \alpha_i^l \in [\alpha_i^-, \alpha_i^+]. \end{cases} \end{aligned} \quad (30)$$

Matrices B_i^l are positive definite and need to be chosen

Algorithm 2 Inexact proximal regularised BCD

Input: Augmented Lagrangian $L_\rho(\cdot, \mu)$, polytopes $\{\mathcal{Z}_i\}_{i=1}^N$ and termination tolerance $\tau > 0$.

Initialization: $z_1^0 \in \mathcal{Z}_1, \dots, z_N^0 \in \mathcal{Z}_N$

$l \leftarrow 0$.

while $\|z^{l+1} - z^l\|_\infty > \tau$ **do**

for $i \in \{1, \dots, N\}$ **do**

$z_i^{l+1} \leftarrow \arg\min_{z_i \in \mathcal{Z}_i} \nabla_{z_i} L_\rho(z_1^{l+1}, \dots, z_N^l, \mu)^\top (z_i - z_i^l) + \frac{1}{2} (z_i - z_i^l)^\top (B_i^l + \alpha_i^l I_{n_i}) (z_i - z_i^l)$

end for

$l \leftarrow l + 1$

end while

carefully in order to guarantee convergence of Algorithm 2, as explained in paragraph V-B. Thus, Algorithm 2 consists in solving convex QPs sequentially. Under appropriate assumptions [4], computations can be partly parallelised. Moreover, computational efficacy can be improved by warm-starting.

B. Convergence analysis

When considering the minimisation of functions $f : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_N} \rightarrow \mathbb{R} \cup \{+\infty\}$ satisfying the KL property [5] and having the following structure

$$f(z) = \sum_{i=1}^N f_i(z_i) + P(z_1, \dots, z_N), \quad (31)$$

where f_i are proper lower semicontinuous and P is twice continuously differentiable, two key assumptions need to be satisfied by the iterates z_i^l of the inexact BCD scheme in order to ensure global convergence [1]: a *sufficient decrease* property and the *relative error* condition. The *sufficient decrease* property asserts that

$$\forall i \in \{1, \dots, N\}, \forall l \in \mathbb{N},$$

$$\begin{aligned} f_i(z_i^{l+1}) + P(z_1^{l+1}, \dots, z_i^{l+1}, \dots, z_N^l) + \frac{\alpha_i^l}{2} \|z_i^{l+1} - z_i^l\|_2^2 \\ \leq f_i(z_i^l) + P(z_1^{l+1}, \dots, z_{i-1}^{l+1}, z_i^l, \dots, z_N^l). \end{aligned} \quad (32)$$

The *relative error* condition states that

$$\forall i \in \{1, \dots, N\}, \exists b_i > 0 \text{ such that } \forall l \in \mathbb{N},$$

$$\exists v_i^{l+1} \in \partial f_i(z_i^{l+1}),$$

$$\begin{aligned} \|v_i^{l+1} + \nabla_{z_i} P(z_1^{l+1}, \dots, z_i^{l+1}, z_{i+1}^l, \dots, z_N^l)\|_2 \\ \leq b_i \|z_i^{l+1} - z_i^l\|_2. \end{aligned} \quad (33)$$

This means that at every iteration one can find a vector in the sub-differential of f_i computed at the current iterate, which is bounded by the momentum of the sequence $z_i^{l+1} - z_i^l$. It

is actually a key step in order to apply the KL inequality, as done in [1]. For clarity, we state Theorem 6.2 in [1].

Theorem 2 (Convergence of inexact BCD): Let f be a proper lower semi-continuous function with structure (31) satisfying the KL property and bounded from below. Let $\underline{\alpha} > 0$ such that $\underline{\alpha} \leq \alpha_i^l$ for all $i \in \{1, \dots, N\}$ and $l \in \mathbb{N}$. Let $\{z^l\}$ be a sequence satisfying (32) and (33). If $\{z^l\}$ is bounded, then $\{z^l\}$ converges to a critical point of f .

Remark 3: Note that $\delta_{\mathcal{Z}} + L_{\rho}(\cdot, \mu)$ has the structure (31).

Our convergence analysis then mainly consists in verifying that the iterates generated by Algorithm 2 satisfy (32) and (33). The following definitions are necessary for the remainder of the proof:

$$\begin{aligned} \forall i \in \{1, \dots, N\}, \forall l \in \mathbb{N}, \\ S(z_1^{l+1}, \dots, z_{i-1}^{l+1}, z_i, z_{i+1}^l, \dots, z_N^l) &:= \\ \mu_G^\top G(z_1^{l+1}, \dots, z_{i-1}^{l+1}, z_i, z_{i+1}^l, \dots, z_N^l) &+ \\ + \frac{\rho}{2} \|G(z_1^{l+1}, \dots, z_{i-1}^{l+1}, z_i, z_{i+1}^l, \dots, z_N^l)\|_2^2, \\ R_i^l(z_i) &:= Q(z_1^{l+1}, \dots, z_{i-1}^{l+1}, z_i, z_{i+1}^l, \dots, z_N^l) \\ &+ S(z_1^{l+1}, \dots, z_{i-1}^{l+1}, z_i, z_{i+1}^l, \dots, z_N^l), \\ \Psi_i(z_i) &:= \mu_i^\top F_i(z_i) + \frac{\rho}{2} \|F_i(z_i)\|_2^2, \\ L_i^l(z_i) &:= J_i(z_i) + \Psi_i(z_i) + R_i^l(z_i), \\ \Gamma_i(z_i) &:= J_i(z_i) + \Psi_i(z_i) + \delta_{\mathcal{Z}_i}(z_i), \\ \Phi_i^l(z_i) &:= L_i^l(z_i) + \delta_{\mathcal{Z}_i}(z_i), \end{aligned} \quad (34)$$

where $\mu_G \in \mathbb{R}^r$ is the subvector of multipliers associated with the equality constraint $G(z_1, \dots, z_N) = 0$.

Lemma 5 (Bound on hessian norm):

$$\forall i \in \{1, \dots, N\}, \forall l \in \mathbb{N}, \quad \|\nabla^2 L_i^l\|_{\infty}^{\mathcal{Z}_i} \leq C_i, \quad (35)$$

where

$$C_i := \|\nabla^2 J_i\|_{\infty}^{\mathcal{Z}_i} + \|\nabla^2 \Psi_i\|_{\infty}^{\mathcal{Z}_i} + \max_{i \in \{1, \dots, N\}} \|\nabla_i^2 (S + Q)\|_{\infty}^{\mathcal{Z}}. \quad (36)$$

Proof: The proof directly follows from the definitions (34) and the fact that the functions involved in (34) are twice continuously differentiable over compact sets \mathcal{Z}_i . ■

Assumption 3: The matrices B_i^l can be chosen so that $B_i^l - C_i I_{n_i} \succ 0$ and $2C_i I_{n_i} - B_i^l \succ 0$, for all $i \in \{1, \dots, N\}$ and all $l \in \mathbb{N}$.

Remark 4: Such an assumption requires knowledge of the Lipschitz constant of the gradient of the augmented Lagrangian. However, a simple backtracking procedure may be applied in practice.

Lemma 6 (Sufficient decrease): For all $i \in \{1, \dots, N\}$ and all $l \in \mathbb{N}$,

$$\Phi_i^l(z_i^{l+1}) + \frac{\alpha_i^l}{2} \|z_i^{l+1} - z_i^l\|_2^2 \leq \Phi_i^l(z_i^l). \quad (37)$$

Proof: By definition of z_i^{l+1} in Algorithm 2,

$$\begin{aligned} \forall z_i \in \mathcal{Z}_i, \\ \nabla L_i^l(z_i^l)^\top (z_i^{l+1} - z_i^l) \\ + \frac{1}{2} (z_i^{l+1} - z_i^l)^\top (B_i^l + \alpha_i^l I_{n_i}) (z_i^{l+1} - z_i^l) \\ \leq \nabla L_i^l(z_i^l)^\top (z_i - z_i^l) + \frac{1}{2} (z_i - z_i^l)^\top (B_i^l + \alpha_i^l I_{n_i}) (z_i - z_i^l), \end{aligned} \quad (38)$$

which by substituting z_i with $z_i^l \in \mathcal{Z}_i$ implies that

$$\begin{aligned} \nabla L_i^l(z_i^l)^\top (z_i^{l+1} - z_i^l) \\ + \frac{1}{2} (z_i^{l+1} - z_i^l)^\top (B_i^l + \alpha_i^l I_{n_i}) (z_i^{l+1} - z_i^l) \leq 0. \end{aligned} \quad (39)$$

However, by applying the descent Lemma, one gets

$$L_i^l(z_i^{l+1}) \leq L_i^l(z_i^l) + \nabla L_i^l(z_i^l)^\top (z_i^{l+1} - z_i^l) + \frac{C_i}{2} \|z_i^{l+1} - z_i^l\|_2^2. \quad (40)$$

Combining this last inequality with inequality (39), one obtains

$$\begin{aligned} L_i^l(z_i^{l+1}) \leq L_i^l(z_i^l) \\ + \frac{1}{2} (z_i^{l+1} - z_i^l)^\top ((C_i - \alpha_i^l) I_{n_i} - B_i^l) (z_i^{l+1} - z_i^l), \end{aligned} \quad (41)$$

which, by Assumption 3, implies that

$$L_i^l(z_i^{l+1}) + \frac{\alpha_i^l}{2} \|z_i^{l+1} - z_i^l\|_2^2 \leq L_i^l(z_i^l). \quad (42)$$

However, $\delta_{\mathcal{Z}_i}(z_i^{l+1}) = \delta_{\mathcal{Z}_i}(z_i^l) = 0$. Thus,

$$\Phi_i^l(z_i^{l+1}) + \frac{\alpha_i^l}{2} \|z_i^{l+1} - z_i^l\|_2^2 \leq \Phi_i^l(z_i^l). \quad (43)$$

■

Lemma 7 (Relative error condition):

$$\begin{aligned} \forall i \in \{1, \dots, N\}, \forall l \in \mathbb{N}, \\ \exists v_i^{l+1} \in \partial \Gamma_i(z_i^{l+1}), \\ \|v_i^{l+1} + \nabla(Q + S)(z_1^{l+1}, \dots, z_i^{l+1}, z_{i+1}^l, \dots, z_N^l)\|_2 \\ \leq b_i \|z_i^{l+1} - z_i^l\|_2, \end{aligned} \quad (44)$$

where $b_i := 3C_i + \alpha_i^+$ and α_i^+ is defined in (30).

Proof: Writing the necessary optimality conditions for the i^{th} QP of Algorithm 2, one obtains

$$0 \in \nabla L_i^l(z_i^l) + (B_i^l + \alpha_i^l I_{n_i})(z_i^{l+1} - z_i^l) + \mathcal{N}_{\mathcal{Z}_i}(z_i^{l+1}). \quad (45)$$

Thus there exists $w_i^{l+1} \in \mathcal{N}_{\mathcal{Z}_i}(z_i^{l+1})$ such that

$$\begin{aligned} 0 = w_i^{l+1} + \nabla L_i^l(z_i^l) + (B_i^l + \alpha_i^l I_{n_i})(z_i^{l+1} - z_i^l) \\ = w_i^{l+1} + \nabla L_i^l(z_i^{l+1}) + \nabla L_i^l(z_i^l) - \nabla L_i^l(z_i^{l+1}) \\ + (B_i^l + \alpha_i^l I_{n_i})(z_i^{l+1} - z_i^l). \end{aligned} \quad (46)$$

The last equality yields

$$\begin{aligned} w_i^{l+1} + \nabla J_i(z_i^{l+1}) + \nabla F_i(z_i^{l+1})^\top \mu \\ + \rho \nabla F_i(z_i^{l+1})^\top F_i(z_i^{l+1}) \\ + \nabla_i(Q + S)(z_1^{l+1}, \dots, z_i^{l+1}, z_{i+1}^l, \dots, z_N^l) \\ = \nabla L_i^l(z_i^{l+1}) - \nabla L_i^l(z_i^l) + (B_i^l + \alpha_i^l I_{n_i})(z_i^l - z_i^{l+1}), \end{aligned} \quad (47)$$

from which it immediately follows

$$\begin{aligned} \exists v_i^{l+1} \in \partial \Gamma_i(z_i^{l+1}), \\ v_i^{l+1} + \nabla_i(Q + S)(z_1^{l+1}, \dots, z_{i-1}^{l+1}, z_i^{l+1}, z_{i+1}^{l+1}, \dots, z_N^{l+1}) \\ = \nabla L_i^l(z_i^{l+1}) - \nabla L_i^l(z_i^l) + (B_i^l + \alpha_i^l I_{n_i})(z_i^l - z_i^{l+1}). \end{aligned} \quad (48)$$

However, as $L_i^l \in C^2(\mathcal{Z}_i, \mathbb{R})$, one gets

$$\begin{aligned} \|v_i^{l+1} + \nabla_i(Q + S)(z_1^{l+1}, \dots, z_{i-1}^{l+1}, z_i^{l+1}, z_{i+1}^{l+1}, \dots, z_N^{l+1})\|_2 \\ \leq (C_i + \|B_i^l\|_2 + \alpha_i^l) \|z_i^{l+1} - z_i^l\|_2 \\ \leq (3C_i + \alpha_i^+) \|z_i^{l+1} - z_i^l\|_2. \end{aligned} \quad (49)$$

Now that the two main ingredients are proven, one can state the theorem guaranteeing convergence of the iterates of Algorithm 2 to a critical point of $\delta_{\mathcal{Z}} + L_{\rho}(\cdot, \mu)$.

Theorem 3 (Convergence of Algorithm 2): The sequence $\{z^l\}$ generated by Algorithm 2 converges to a point z^* satisfying

$$0 \in \nabla L_{\rho}(z^*, \mu) + \mathcal{N}_{\mathcal{Z}}(z^*) \quad (50)$$

or equivalently $d(0, \nabla L_{\rho}(z^*, \mu) + \mathcal{N}_{\mathcal{Z}}(z^*)) = 0$.

Proof: The sequence $\{z^l\}$ is obviously bounded, as the iterates are constrained to stay in the polytope \mathcal{Z} . The function $\delta_{\mathcal{Z}} + L_{\rho}(\cdot, \mu)$ is bounded from below. Moreover, by Assumption 1, $\delta_{\mathcal{Z}} + L_{\rho}(\cdot, \mu)$ satisfies the KL property for all $\mu \in \mathbb{R}^m$ and $\rho > 0$. The sequence $\{z^l\}$ satisfies the conditions (32) and (33). The parameter $\underline{\alpha}$ of Theorem 2 can be taken as $\min \{\alpha_i^- | i \in \{1, \dots, N\}\}$. Convergence to a critical point z^* of $\delta_{\mathcal{Z}} + L_{\rho}(\cdot, \mu)$ then follows by applying Theorem 2. ■

We can then guarantee that, given $\epsilon > 0$, by taking a sufficiently large number of inner iterations, Algorithm 2 converges to a point $\bar{z} \in \mathcal{Z}$ satisfying $d(0, \nabla L_{\rho}(\bar{z}, \mu) + \mathcal{N}_{\mathcal{Z}}(\bar{z})) \leq \epsilon$, as needed by Algorithm 1.

VI. NUMERICAL EXAMPLE

Our algorithm is tested on the following class of non-convex NLPs:

$$\begin{aligned} \text{minimise}_{x_1, \dots, x_N} \sum_{i=1}^N x_i^T H_i x_i + \sum_{i=1}^{N-1} x_i^T H_{i,i+1} x_{i+1} \\ \text{s.t.} \\ \|x_i\|_2^2 = a^2, \quad i \in \{1, \dots, N\} \\ -b \leq x_{i,j} \leq b, \quad i \in \{1, \dots, N\}, j \in \{1, \dots, d\}, \end{aligned} \quad (51)$$

where N is the number of agents and d their dimension. Matrices H_i and $H_{i,i+1}$ are indefinite. When applying Algorithm 2 to minimise the augmented Lagrangian associated with (51), the update of agent i only depends on agents $i-1$ and $i+1$, which implies that, after appropriate re-ordering [4], every inner iteration actually consists of two sequential update steps where $N/2$ QPs are solved in parallel. We fix $d = 3$, $N = 20$, $a = \sqrt{R}$ and $b = 0.6R$ with $R = 2$. Matrices H_i and $H_{i,i+1}$ are randomly generated. For a fixed number of outer iterations, we gradually increase the number of inner iterations and count the number of random problems, on which the algorithm ends up within a fixed feasibility

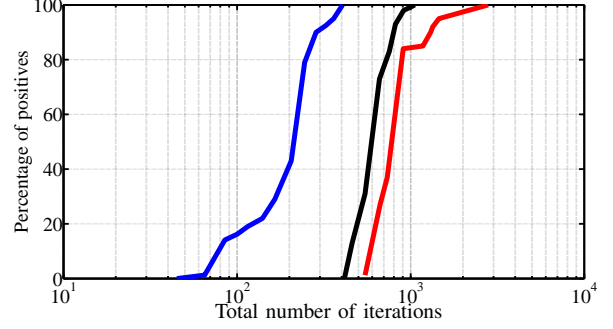


Fig. 1. Percentage of positive problems for a fixed feasibility tolerance on the nonlinear equality constraints. Feasibility margins 10^{-3} in blue, 10^{-4} in black and 10^{-6} in red.

tolerance with respect to the nonlinear equality constraints $\|x_i\|_2^2 = a^2$. Statistics are reported for 500 random NLPs of the type (51). The initial penalty parameter is set to $\rho_0 = 0.1$ and the growth coefficient $\beta = 100$. The hessian matrices of Algorithm 2 are set to $30\rho I_d$. The algorithm is initialised on random primal and dual initial guesses, feasible with respect to the inequality constraints. Results for a feasibility tolerances of 10^{-3} , 10^{-4} and 10^{-6} are presented in Figure 1. It clearly appears that obtaining an accurate feasibility margin requires a large number of iterations, yet a reasonable feasibility (10^{-3}) can be obtained with approximately hundred iterations.

REFERENCES

- [1] H. Attouch, J. Bolte, and B.F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting and regularised Gauss-Seidel methods. *Mathematical Programming*, 137:91–129, 2013.
- [2] D.P. Bertsekas. Convexification procedures and decomposition methods for nonconvex optimisation problems. *Journal of Optimization Theory and Applications*, 29:169–197, 1979.
- [3] D.P. Bertsekas. *Constrained optimisation and Lagrange multiplier methods*. Athena Scientific, 1982.
- [4] D.P. Bertsekas and J.N. Tsitsiklis. *Parallel and distributed computation: numerical methods*. Athena Scientific, Belmont, MA, 1997.
- [5] J. Bolte, A. Daniilidis, and A. Lewis. The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17:1205–1223, 2006.
- [6] A. R. Conn, G. I. M. Gould, and P. L. Toint. A globally convergent augmented Lagrangian algorithm for optimisation with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28(2):545–572, 1991.
- [7] A. Hamdi and S.K. Mishra. Decomposition methods based on augmented Lagrangian: a survey. In *Topics in nonconvex optimization*. Mishra, S.K., 2011.
- [8] I. Necoara, V. Nedelcu, and I. Dumitrache. Parallel and distributed optimization methods for optimisation and control in networks. *Journal of Process Control*, 21:756–766, 2011.
- [9] I. Necoara, C. Savorgnan, Q. Tran Dinh, J. Suykens, and M. Diehl. Distributed nonlinear optimal control using sequential convex programming and smoothing techniques. In *Proceedings of the 48th Conference on Decision and Control*, 2009.
- [10] R.T. Rockafellar. Augmented Lagrangian multiplier functions and duality in nonconvex programming. *SIAM Journal on Control and Optimization*, 12(2):268–285, 1974.
- [11] R.T. Rockafellar and R. Wets. *Variational Analysis*. Springer, 1998.
- [12] A. Tanikawa and H. Mukai. A new technique for nonconvex primal-dual decomposition of a large-scale separable optimisation problem. *IEEE Transactions on Automatic Control*, 1985.